ORIGINAL

NATIONAL UNIVERSITY OF SINGAPORE

EXAMINATION FOR

(Semester I: 2019/2020)

EE5731 – VISUAL COMPUTING

November / December 2019 - Time Allowed: 2.5 Hours

INSTRUCTIONS TO CANDIDATES:

1. This paper contains **Seven (7)** questions and comprises **FOUR (4)** printed pages.
2. Answer all questions and answer them as detailed as possible. Give the impression that you really understand the answers, as it is critical part of the marking.
3. This is a CLOSED BOOK examination but candidates are allowed to bring in **ONE (1)** A4 sheet of paper, on which they may write any information, into the examination hall.
4. Total marks are 100.

Q1: Referring to **Figure Q1**, by modifying $\beta = \frac{2\epsilon_t}{1-2\epsilon_t}$, what are the impacts on the whole training process and the face detection results? Provide the justification of your answer.

[15 points]

- For $t = 1, \ldots, T$:
  1. Normalize the weights, $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$
  2. Select the best weak classifier with respect to the weighted error
  $$\epsilon_t = \min_{f,p,\theta} \sum_i w_i \, | h(x_i, f, p, \theta) - y_i | .$$
  3. Define $h_t(x) = h(x, f_t, p_t, \theta_t)$ where f_t, p_t, and θ_t are the minimizers of ϵ_t.
  4. Update the weights:
  $$w_{t+1,i} = w_{t,i}\beta_t^{1-e_i}$$
  where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- The final strong classifier is:
$$C(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$
where $\alpha_t = \log \frac{1}{\beta_t}$

Figure Q1: Part of Viola-Jones algorithm for face detection

Q2: SIFT (Scale-Invariant Feature Transform) is useful for many applications, since it can provide features and descriptors that are robust to some changes including scale, lighting, rotation, etc. Considering **Figure Q2.B**, where the right image is a mirrored version of the left image, will SIFT be also invariant to this change? Justify your answer with respect to all the 4 steps of SIFT.

Figure Q2.B: These two images, (a) and (b), are mirrored to each other

[15 points]

Q3: Regarding depth from video, the geometric coherence constraint can be expressed as:

$$\rho_v(\bar{x}, d, D_{t'}) = \exp\left(-\frac{||\bar{x} - l_{t',t}(\bar{x}', D_t(\bar{x}'))||}{2\sigma_d^2}\right).$$

A. What is the geometrical meaning of every variable in the equation?
B. How does the geometric coherence constraint work? Discuss in detail and draw a diagram illustrating the geometric coherence constraint.

[15 points]

Q4: We can define $\boldsymbol{F} = [e']_\times \boldsymbol{H}_\pi$, where $\boldsymbol{F}$ is the fundamental matrix and $\boldsymbol{H}_\pi$ is the homography matrix, and e' is the epipole. Since $\boldsymbol{H}_\pi$ is dependent on the scene (namely, $x' = \boldsymbol{H}_\pi x$ is true, only if x and x' refers to the same point on the plane in the 3D world as shown in **Figure Q3.B**), does it mean that $\boldsymbol{F}$ is also dependent on the scene? Justify your answer.

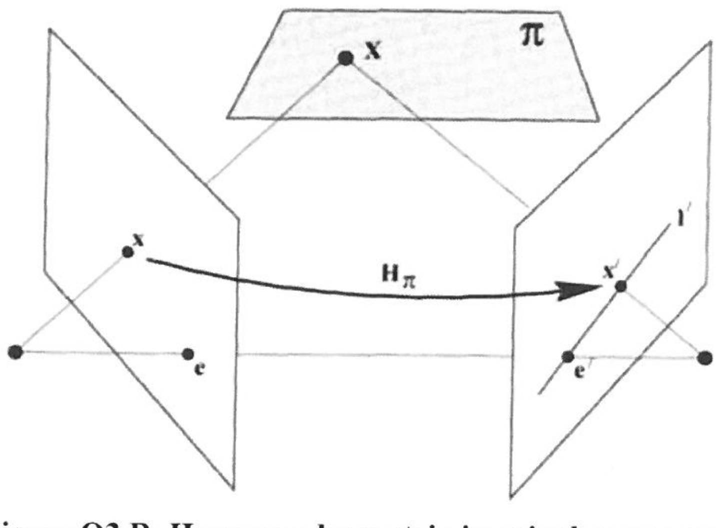

Figure Q3.B: Homography matrix in epipolar geometry

[15 points]

Q5: Semantic segmentation is a task to segment an image into semantic classes. An example of semantic segmentation is shown in **Figure Q4**, where every pixel is labeled according to its class (i.e. either sky, water, forest, grass, etc.). Suppose you are given a task to solve this semantic segmentation using Markov Random Field (MRF), write your step-by-step algorithm (including the definition of your data and prior terms) based on the knowledge we learn in the course.

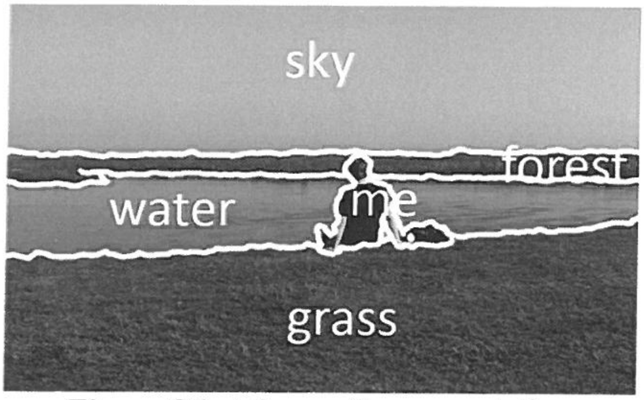

Figure Q4: A Semantic segmentation

[15 points]

Q6: Optical flow and depth from stereo are related to some extent. Assuming that the input image pair for optical flow and depth from stereo is the same image pair, answer the following questions:

A. Explain how to compute optical flow from the depth obtained from stereo.
B. Mention 2 conditions that make computing depth from optical flow fail. Justify your answer.

[15 points]

Q7: In structure-texture decomposition using ROF (Rodin-Osher-Fatemi), the error function can be expressed as:

$$E(I_s(\bar{x})) = (I_s(\bar{x}) - I(\bar{x}))^2 + \lambda||\nabla I_s(\bar{x})||_2.$$

If we optimize this error function globally, we will obtain a significantly less noisy and cartoon-like structure layer. Explain why the total variation (TV) loss can generate a less noisy and cartoon-like structure, while at the same time generate sharper edges (sharper than using, e.g., the quadratic L2-norm).

[10 points]

END OF PAPER